

**Problem 1**

To enforce the necessary constraints, we introduce the Lagrangian multipliers  $\lambda_1, \dots, \lambda_M, \lambda_{M+1}$  and make an unconstrained maximization of

$$L = \mathbf{u}_{M+1}^\top \mathbf{S} \mathbf{u}_{M+1} + \lambda_{M+1}(1 - \mathbf{u}_{M+1}^\top \mathbf{u}_{M+1}) + \sum_{i=1}^M \lambda_i (\mathbf{u}_{M+1}^\top \mathbf{u}_i)$$

Now we take the derivative of  $L$  with respect to a vector  $\mathbf{u}_{M+1}$  and set it to zero

$$\begin{aligned} \frac{\partial}{\partial \mathbf{u}_{M+1}} L = 0 &\Leftrightarrow 2\mathbf{S} \mathbf{u}_{M+1} = 2\lambda_{M+1} \mathbf{u}_{M+1} - \sum_{i=1}^M \lambda_i \mathbf{u}_i \\ &\Leftrightarrow 2\mathbf{u}_{M+1}^\top \mathbf{S} \mathbf{u}_{M+1} = 2\lambda_{M+1} \mathbf{u}_{M+1}^\top \mathbf{u}_{M+1} - \sum_{i=1}^M \lambda_i \mathbf{u}_{M+1}^\top \mathbf{u}_i \quad (\text{left-multiply by } \mathbf{u}_{M+1}^\top) \\ &\Leftrightarrow \mathbf{u}_{M+1}^\top \mathbf{S} \mathbf{u}_{M+1} = \lambda_{M+1} \quad (\text{by orthonormality of } \mathbf{u}_i \text{ and } \mathbf{u}_{M+1}) \end{aligned}$$

By induction step, we see that the variance in direction  $\mathbf{u}_{M+1}$  is maximum when we set  $\mathbf{u}_{M+1}$  equal to the eigenvector having the  $(m+1)^{\text{th}}$  largest eigenvalue  $\lambda_{M+1}$ .

**Problem 2**

$$\mathbf{x}_i \sim \mathcal{N}(\mu_{\mathbf{x}}, \Phi_{\mathbf{x}}) \quad (\mu_{\mathbf{x}} = \mu, \Phi_{\mathbf{x}} = \mathbf{W} \mathbf{W}^\top + \Phi^2 \mathbf{I})$$

From lecture we know that  $\mathbf{y}_i = \mathbf{A} \mathbf{x}_i \sim \mathcal{N}(\mu_{\mathbf{y}}, \Phi_{\mathbf{y}})$ . Let's first derive the two moments

$$\begin{aligned} \mu_{\mathbf{y}} &= \mathbf{E}[y] \\ &= \mathbf{E}[\mathbf{A} \mathbf{x}] \\ &= \mathbf{A} \mathbf{E}[\mathbf{x}] \\ &= \mathbf{A} \mu_{\mathbf{x}} \end{aligned}$$

$$\begin{aligned} \Phi_{\mathbf{y}} &= \mathbf{E}[(\mathbf{y} - \mu_{\mathbf{y}})(\mathbf{y} - \mu_{\mathbf{y}})^\top] \\ &= \mathbf{E}[(\mathbf{A} \mathbf{x} - \mathbf{A} \mu_{\mathbf{x}})(\mathbf{A} \mathbf{x} - \mathbf{A} \mu_{\mathbf{x}})^\top] \\ &= \mathbf{E}[(\mathbf{A}(\mathbf{x} - \mu_{\mathbf{x}}))(\mathbf{A}(\mathbf{x} - \mu_{\mathbf{x}}))^\top] \\ &= \mathbf{A} \mathbf{E}[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{x} - \mu_{\mathbf{x}})^\top] \mathbf{A}^\top \\ &= \mathbf{A} \Phi_{\mathbf{x}} \mathbf{A}^\top \end{aligned}$$

By pattern matching we have for the transformed Maximum Likelihood estimates

$$\begin{aligned}\mu_{y_{ML}} &= \mathbf{A}\mu_{ML} \\ \boldsymbol{\Phi}_{y_{ML}} &= \mathbf{A}\boldsymbol{\Phi}_{ML}\mathbf{A}^\top \\ \mathbf{W}_{y_{ML}} &= \mathbf{A}\mathbf{W}_{ML}\end{aligned}$$

By orthogonality  $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top\mathbf{A} = \mathbf{I}$  and  $\boldsymbol{\Phi} = \sigma^2\mathbf{I}$  we have

$$\begin{aligned}\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^\top &= \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^\top \\ &= \sigma^2\mathbf{A}\mathbf{A}^\top \\ &= \sigma^2\mathbf{I}\end{aligned}$$

### Problem 3

Let  $\mathbf{x} \in \mathbb{R}^5$  hold the movie ratings given by Leslie.

By the SVD projection in concept space we have

$$\mathbf{V} \cdot \mathbf{x} = [1.74, 2.84]^\top$$

By SVD decomposition and reconstruction of the input using the projected space we have

$$[1.74, 2.84]^\top \cdot \mathbf{V}^\top = [1.0092, 1.0092, 1.0092, 2.0164, 2.0164]^\top$$

E.g. we can predict that Leslie will rate Titanic movie with 2.0164.

### Problem 4

See below.